SECURITY CLASSIFICATION OF THIS PAGE

AD-A248 479

REPORT DOCUMENTATION PAGE

DTIC ELECTE APR 0 8 1992 S B D

	1b RESTRICTIVE MARKINGS
	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release: distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 92 0191
6a NAME OF PERFORMING ORGANIZATION: Mass. Institute of Technology — 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Res. N/L
6c. ADDRESS (City, State, and ZIP Code) Dept. of Brain & Cognitive Sciences 79 Amherst St./E10-120/Cambridge, MA 02139	7b. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling Air Force Base Washington, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR — 8b OFFICE SYMBOL (If applicable): NL	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR 89-0504
8c. ADDRESS (City, State, and ZIP Code) Same as 7b	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
61102F	2313	A9	

11. TITLE (Include Security Classification)
"Top-Down" Influences on "Bottom-Up" Processing

12. PERSONAL AUTHOR(S) Whitman Richards

13a TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Annual	FROM 10-9-90 TO 10-9-91	4 Feb 1992	7

16. SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Vision, AI, Cognition, Neurophysiology, Visual Psychophysics, Dynamical Systems

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Perception is not simply a bottom-up process, but involves inductive inferences that use "top-down" knowledge to interpret image data. In computational vision, typically this knowledge appears in the form of "constraints" such as rigidity, viewpoint consistency, smoothness, etc. However these constraints are fallible – they do not always apply, and hence the perceptual process entails inductive reasoning. A major theoretical thrust of our work is to provide a formal "lattice" framework for organizing the plausible states of this reasoning aspect of perception (Jepson & Richards). The proposal makes strong predictions, given a set of constraints and a particular picture as to what interpretations or "percepts" will be seen. Consequently we have a series of experiments underway to (1) understand which constraints (or premises)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. John Tangney	22b TELEPHONE (Include Area Code): (202) 767-5021 — 22c OFFICE SYMBOL: NL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

U.S. Government Printing Office: 1985—507-047

ABSTRACT (continued):

are typically invoked in the interpretation of simple line drawings and (2) to show that the "lattice" framework specifies all of these interpretations, placing them in proper rank order.

In parallel, we are also exploring two other models for merging bottom-up and top-down information, both of which are neural-based. One, called sequence-seeking (Ullman), proposes a network heirarchy where a sequence of transformations of both the input data and the target model occur in parallel, searching for the proper mapping that brings each into register. The proposal makes a special effort to incorporate what we currently know about cortical machinery, and also has triggered psychophysical experiments. (We have not yet explored the relations between the lattice and sequence-seeking proposals.)

Finally, there are some studies related to our ability to switch between sets of premises, or to alter our models. These assume "neural-like" states that must interplay with one another. At the moment, we can show that this process is not Poisson, as proposed for very low-level multi-stable percepts, but is more likely characterized as a non-linear, dynamical system.

Annual Report
on
Top-Down Influences on
Bottom-Up Processing
(January 1992)

Whitman Richards
MIT E10-120
79 Amherst St.
Cambridge, MA 02139

617-253-5776
whit@ai.mit.edu

AFOSR 89-0504

As in the previous year, our research effort breaks down roughly into three categories:

1. Theoretical Studies: A Formal Framework for Percepts
   A Neural Proposal for Recognition
2. Experiments related to the above
3. Studies of Dynamical Systems Behavior (Chaos in Percepts)

1. Theoretical Studies

Here we have two main thrusts, one concerned with the logical, formal structure that underlies the act of perception (Richards, Jepson & Feldman) and the other a proposal for how neural machinery might match the incoming sense data to an internal model (Ullman).

1.1 Logic in Percepts (Richards & Jepson)

This work began about two years ago, when we realized that although many are studying "Perception", there is no formal definition of just what a percept is. Without such a definition, how can we decide whether a particular machine or biological state (or model output) qualifies as a perception? Furthermore, how can we build a true theory of a percept without a clear specification of the kinds of state variables, operations, and "language" that are entailed?

Our first answer to "What Is a Percept?" was to note that perceptions are inductive inferences. When conclusions about a state in the world are drawn from the sense data, then (fallible) premises must be proposed to complete the inference process. Because these premises are fallible – they are simply intelligent guesses – a partial order can be placed upon possible interpretations of the sense data, given the chosen premises. The order is determined by ranking the premise combination that must be "given up". Within such an order, a percept can then be defined as a maximal node. (This is not equivalent to minimizing the faulted premises.) The key to locating these maximal nodes is to be able to reason about the consistency of the data, given the current state of "top-down" knowledge (Jepson & Richards, 1991). In a forthcoming paper, "Lattice Framework for Integrating Vision Modules", we compare a specialized version of our proposal to several others, such as probabalistic reasoning and Hough transform schemes that are often used to resolve conflicting conclusions reached by different sense modules. (A simple example of such a conflict would be when you view the TV screen: motion information implies the scene is three dimensional, but your binocular system claims the scene is flat.)

More recently, we have made some major revisions in our "Lattice Theory for Percepts", showing more clearly the structure of the reasoning process involved, as well as a further clarification of the components needed to support a formal theory of perception. Our aim here is to show how the perceptual interpretation process can be cast in terms of first-order logic. Thus, in Richard Gregory's or Irwin Rock's terms, we actually put the logic into percepts, allowing us eventually to run a program that reasons about the sense data (a picture).

1.2 A Neural Proposal (Ullman)

At a completely different level, Ullman has proposed a network heirarchy scheme for how "bottom-up" information comes into register with "top-down" models. The basic process, termed "sequence-seeking", is a search for a sequence of mappings or transformations linking a source and target representation. The search is bidirectional throughout the heirarchy – "bottom-up" as well as "top-down". The novel part of the proposal is that the two searches are performed along two separate, complementary pathways, one ascending, the other descending. When a matching pattern is found, regardless of the level, then a chain of activity linking the source and target is generated, facilitating one particular path in the network. The proposal is largely consistent with what is known about cortical machinery, specifically the interplay between the various visual areas, and hence is a hypothesis about the basic scheme of information processing in the neocortex (and thalamus). Experiments related to this proposal are currently underway – see below.

1.3 "Good" Features and Categories (Richards, Jepson & Feldman)

Certain image properties, such as collinearity, parallel lines, V-junctions, allow strong inferences about the 3D configuration that creates these image features (Binford, Lowe). Other image projectons, such as "X", do not generally support strong inferences because either they occur generically in the image, or because they can arise in many different ways, as a T-junction can. We now have completed a specification for just what constitutes a "good" feature – i.e. one with strong inferential powers. In this specification, we show how a measure can be placed on the inductive power of a feature, namely the codimension of the arrangement in the chosen model class (see Richards & Jepson, 1992). This work is important for two reasons: (1) it tells us which image features are worth computing and (2) these features then, in turn, dictate just how a "feature space" should be subdivided to form useful categories within that space. (See Richards & Jepson, 1992, for "Feature" analysis, and Feldman, 1992, for preliminary work on "Categories as Subspaces".) Curiously, it turns out that ideas from catastrophy theory (Poston & Stewart) and solid shape (Koenderink) provide some of the formal structure for both the feature and category work. This is because a feature space can be considered to be a surface in a particular manifold, and hence the problem reduces in part to when such surfaces should be distinguished and how coordinate frames should be assigned to them (such as choosing principal curvatures). [Note: a longer-term project related to categories and conceptual theories – such as in physics – is underway with J.J. Koenderink.]

2.0 Experiments

These will be subdivided roughly into sections similar to the above.

2.1 Lattice Theory for Percepts (Richards & Jepson)

Our principal experimental result over the past year is a study of a rigid 3D configuration that appears non-rigid: the rotating Ames Trapezoid window plus stick. When seen in kinetic depth, such as on a TV screen, the motion of the configuration satisfies all theories (such as Ullman's, Hoffman & Bennett, Huang, etc.) that predict rigid motion (indeed, the motion is simulated to arise from a rigid object). Yet everyone sees the bar moving non-rigidly with respect to the window. Our "Lattice" explanation for this percept is given in the attached SMC paper: "A Lattice Framework for Integrating Vision Modules". As proposed above in Section 1, our answer is simply that, given very reasonable premises about structures in the world, a non-rigid interpretation will be preferred. (A key to the explanation is that part-decompostions are a necessary step in image understanding, and once we have committed to "parts", the space of plausible solutions becomes very restricted.)

We have also completed an analysis of another deceptively complex configuration – a simple triangle with a stick crossing one of its sides. The stick typically appears to be oriented in space like a handle, with its end-point lying in the plane of the triangle. Why should the stick appear to have the angle it does, and why does the endpoint touch the triangle rather than free-float in space? Again, very simple premises generate a lattice that answers these questions and accounts for our percepts. (We are in the process of writing up this study.)

In addition to the above, we also have another handful of simple displays under analysis. The aim here is to see if the *same* collection of premises can be used to explain the percepts seen for quite different displays. Our current emphasis is on premises dealing with how objects are supported.

2.2 Neural Mechanisms for Recognition

In contrast to the previous studies that address cognitive issues, the following experiments are aimed at understanding neural machinery.

2.2.1 Configuration Stereopsis (Richards)

We are just winding up a study on 3D shape that relates to how "top-down" information about fixation distance (or shape) modulates angular disparity. Because binocular disparity appears to be computed in V2, this modulation *must* occur early in the visual pathway and hence is potentially accessible to psychophysical probing.

As the distance to an object increases, the angular disparity needed to measure the actual 3D configuration must decrease (reaching zero at the horizon). However, if we take an object, say a cup, and evaluate its 3D shape nearby versus far away, the cup does not appear to flatten, although the disparity signal becomes much smaller as the distance increases. This suggests a rescaling of disparity with object (or fixation) distance.

We have conducted parametric studies of 3D shape from stereo over a wide range of fixation distances. The data show that indeed, the depth measure associated with a fixed angular disparity changes with fixation distance. The effect is in the direction needed to preserve the shape of 3D configurations as their distance changes, and is roughly two-thirds of what is needed for a full correction. This is evidence for neural signals being modified at or before the extraction of binocular disparity. Hence we have a preliminary "handle" on how a simple case of "bottom-up" information – namely binocular disparity – may incorporate a form of "top-down" knowledge.

2.2.2 Texture Curvature (with Hugh Wilson)

This study examines curvature discrimination for edges created by texture contours, and includes a model incorporating end-stopped complex cells. (The manuscript is under review by JOSA A (copy enclosed).)

2.3 Computational Vison

We have two studies now underway in this area, one on shape-from-shading and stereo, the other just beginning on color texture.

2.3.1 Shading and Stereo (Dawson & Shashua)

Pseudo stereopsis is when the binocular disparities of a surface, such as a face, are reversed but the shading is not. The impression is that the face is "normal" – the nose, for example, still points outward to the viewer.

We have manipulated noses using graphics techniques in order to push them inward, "into the head" so to speak, without altering the shading. No one is able to see these noses "shoved in". Our analysis suggests that this failure of stereopsis is simply due to the shape-from-shading solution "overriding" (in the Percepts Lattice sense) the weak stereo signal created by shaded rather than sharp contours. The effect is not special to faces, and occurs also for "playdo" shapes.

2.3.2 Color Texture (with D.D. Hoffman et al. at Irvine)

Although much work has been devoted to understanding the appearance of homogenous color patches, almost nothing is known about how we represent colored textures. Our approach is to consider the spatial texture pattern as generated by a Markovian process, which "paints" different colors on a surface. The problem, then, is to recover the characteristic parameters of this underlying process.

This problem is almost ideally suited to the formalism described in Observer Mechanics (Bennett, Hoffman & Prakash), because Markovian kernels lie at the heart of this theory. On the experimental side, we know from earlier work on "Texture Matching" that there will be severe psychophysical restrictions on discriminable patterns, just like in color matching, and expect to find further constraints imposed upon color-texture matches. (Julesz studied this briefly many years ago.)

To date, we have met for three days on this problem at Irvine.

3.0 Dynamical Systems (Chaos in Percepts)

This project continues to advance, although slowly. The underlying hypothesis is that switching between multi-stable percepts, or switching between premises, must be rapid. A chaotic-like, dynamical system would be an attractive mechanism.

Finally (!) we have been able to prove experimentally that very high level perceptual processing is chaotic, with a dimension roughly 3. Our difficulty has been in getting data that is sufficiently free from noise. We have such data now and the result is clear.

Our next step is to develop a model for the effects of noise on a non-linear time series process (with Jepson) and to show that this model applies to our earlier, noisy data. We then will be able to extract the underlying noise-free dynamical process and can proceed to model this process.

4.0 Publications (to date)

Jepson, A. & Richards, W. (1991) What is a percept? Cognitive Science Paper #43, Center for Cognitive Science, MIT.

Jepson, A. & Richards, W. (1992) A lattice framework for integrating modules. *IEEE-SMC,* in press.

Moses, Y., Schechtman, G. & Ullman, S. (1990) *Self-calibrated collinearity detector. Biol. Cybern.*, **63**: 463-475.

Wilson, H. & Richards, W. (1992) Curvature Discrimination at Texture Boundaries. (Under review.)

Feldman, J. (1991) Perceptual simplicity and modes of structural generation. *Proc. Cog. Soc.*

Richards, W. & Jepson, A. (1992) What makes a good feature? In (eds) M. Jenkin & I. Howard, *Spatial Vision in Humans and Robots*, in press.

Ullman, S. (1992) Sequence seeking and counter streams. *MIT AI Memo 1311.*

Feldman, J., Epstein, D. & Richards, W. (1992) Force dynamics of tempo change in music. *Music Perception*, in press.

Koenderink, J.J. & Richards, W. (1992) Why Is Snow So Bright? *Jrl. Opt. Soc. Am. A*, in press.

Bennett, B., Hoffman, D.D & Richards, W. (1991) Reasoning under uncertainty: Lebesgue logic and order logic. *University of California, Irvine, Math. Beh. Sci. Memo MBS 91-08.*

In Preparation:

Logic in percepts (with A. Jepson).

From features to categories (with A. Jepson & J. Feldman).

Configuration stereopsis (W. Richards).

Choosing a coordinate frame (with J. Brian Subirana-Vilanova). (See "Figure-ground in visual perception" ARVO 1991 for brief presentation.

Shading and stereo (with B. Dawson & A. Shashua).

Talks:

University of Minnesota (May 1989) "'Perception and perceivers".

Harvard University (Nov. 1989) "What's a perception?"

Yale University (May 1990) "What's a percept?"

University of Michigan (June 1990) "What's a percept?"

Cognitive Science Society (July 1990) "Perception, computation and categorization".

Cornell University (June 1991) "What makes a good feature?"

York University (June 1991) "Integrating vision modules".

University of Illinois (Oct. 1991) (1) "What's a percept?, (2) "Choosing coordinate frames".

5.0 Funds and Personnel

At the end of the second fiscal year (Oct. 1991) we were over budget by $2,600. We expect to reocver this overrun in 1992, due to the absence of Shimon Ullman, who is at Weisman, but who will return for two months this summer (June and July 1992). We also expect to support in part Jacob Feldman over part of this year.